Received: 3 February 2023 | Revised: 28 June 2023 | Accepted: 22 August 2023

DOI: 10.1002/psp4.13054

ARTICLE

Integrating machine learning with pharmacokinetic models: Benefits of scientific machine learning in adding neural networks components to existing PK models

Diego Valderrama[1,2] | **Ana Victoria Ponce-Bobadilla**[3] | **Sven Mensing**[3] | **Holger Fröhlich**[1,2] | **Sven Stodtmann**[3]

[1]Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Sankt Augustin, Germany

[2]Bonn-Aachen International Center for Information Technology (B-IT), University of Bonn, Bonn, Germany

[3]AbbVie Deutschland GmbH & Co. KG, Ludwigshafen, Germany

Correspondence
Holger Fröhlich, Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, 1, Sankt Augustin 53757, Germany.
Email: holger.froehlich@scai.fraunhofer.de

Sven Stodtmann, AbbVie Deutschland GmbH & Co. KG, Knollstrasse 50, Ludwigshafen am Rhein 67061, Germany.
Email: sven.stodtmann@abbvie.com

Abstract

Recently, the use of machine-learning (ML) models for pharmacokinetic (PK) modeling has grown significantly. Although most of the current approaches use ML techniques as black boxes, there are only a few that have proposed interpretable architectures which integrate mechanistic knowledge. In this work, we use as the test case a one-compartment PK model using a scientific machine learning (SciML) framework and consider learning an unknown absorption using neural networks, while simultaneously estimating other parameters of drug distribution and elimination. We generate simulated data with different sampling strategies to show that our model can accurately predict concentrations in extrapolation tasks, including new dosing regimens with different sparsity levels, and produce reliable forecasts even for new patients. By using a scenario of fitting PK data with complex absorption, we demonstrate that including known physiological structure into an SciML model allows us to obtain highly accurate predictions while preserving the interpretability of classical compartmental models.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

Frameworks that combine mechanistic and machine learning tools (scientific ML [SciML] approaches) have shown promising results in pharmacokinetic (PK) modeling. Methods that allow to capture and parametrize known and unknown mechanisms have not been explored widely.

WHAT QUESTION DID THIS STUDY ADDRESS?

Can PK models be enhanced by adding neural network terms to capture unknown mechanisms? Can these frameworks make accurate PK predictions and how does its performance compared to other SciML approaches?

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

The proposed PK-SciML model learns an unknown absorption mechanism and PK relevant parameters simultaneously. This framework does not require large

or dense sampling datasets to accurately perform typical extrapolation tasks faced during drug development.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?

Our approach sets the scene for developing understandable and reliable approaches for many applications in the PK/pharmacometric field. The method has many desirable properties and can be used as stepping-stone for further method development.

INTRODUCTION

Machine-learning (ML) applications in clinical pharmacology have been rapidly increasing over the last several years. Methods in this area have been focused on (but not limited to) data imputation,[1] covariate patient stratification,[2–5] and predictive modeling.[6–10] Different reviews[11,12] have surveyed the current methods available and described considerations for future methods, such as interpretability, generalizability, and reproducibility.

This work focuses on predictive models for pharmacokinetics (PKs). Most of the existing neural network approaches do not incorporate physiological structure into the model,[13,14] however, more recently, some works have started to do this.[6,9] For example, work conducted by Qian et al.[6] considers a neural network architecture which takes into consideration an ordinary differential equation (ODE) system of the PK dynamics. The work of Janssen et al.[9] considers a physiologically based and well-known compartmental model for describing PK concentrations and uses a neural network term for learning covariate effects, which are incorporated as ODE parameters. Combining mechanistic frameworks, specifically differential equations, and ML – (called scientific machine learning [SciML]),[15,16] allows us to benefit from the advantages of both methodologies. The mechanistic framework allows for the inclusion of physiological constraints and domain expertise in a set of equations, whereas an ML architecture helps manage multimodal (and potentially high dimensional) data, unknown mechanisms, and missing values. The use of SciML models in other fields of science have yielded encouraging results[17,18]; thus, providing the motivation for this work to apply a similar strategy to clinical pharmacology.[15,16,19] We propose an extension of the typical pharmacometrics workflow, where an initial compartmental model is fitted to a dataset and a trained pharmacometrician identifies shortcomings of the model. It is in this step that we introduce the PK-SciML framework, which builds on an existing compartmental PK model and couples it with one or more neural network terms to describe unknown mechanisms, such as absorption or clearance. In the main text of this paper, we will focus on a complex absorption as an unknown mechanism, but the principle can be applied to any component of the compartmental model.

Our reasons to do this are two-fold. We want to enable the estimation of "known-unknowns," such as clearance, or volume of distribution along with a flexible neural network term to enhance interpretability of the results. A second motivation is that we hypothesize that by adding known physiological structure, less data will be needed to train the model compared to a purely data driven approach and that the typical extrapolation tasks asked from a PK model will yield more robust results.

To test these hypotheses, the performance of the proposed method in typical scenarios, that are frequently considered for population PK models, is compared against two established purely data driven architectures.[7,8] The first by Lu et al.,[7] is based on the neural ODE approach,[13] in which the initial conditions of an ODE system parameterized via a neural network are learned via a gated-recurrent unit (GRU) neural network from data. The second model, by Bräm et al.,[8] is based solely on a recurrent neural network architecture. These scenarios are important for performance assessments because the questions in PKs clearly go beyond predictions for samples similar to those in the training set: more specifically, relevant tasks include extrapolation beyond the time horizon provided in the data, prediction of compartmental concentrations for new doses, predictions for missing doses, and prediction in case of a complete dosing cessation.

As an initial example, we consider the case of a complex absorption model. Capturing absorption mechanisms appropriately is a hurdle even for state-of-the-art parameter estimation methods like the nonlinear mixed effects (NLME) model.[20–22] Different works have highlighted the importance and the difficulty of modeling drug absorption profiles.[20–22] We show that neural networks can be used to characterize these absorption profiles. Furthermore, we demonstrate that they aid a compartmental model to capture the behavior of the absorption mechanism while maintaining interpretability and physiological relevance, allowing for robust extrapolation.

METHODS

Problem definition

In this paper, we consider the problem of fitting PK data of a specific drug that exhibits a potentially atypical absorption process. We propose a model that learns an unknown realistic absorption mechanism using neural networks and estimates PK relevant parameters simultaneously. We compare this model against state-of-the-art architectures recently proposed for similar tasks in several evaluation scenarios.

Training data and scenarios

Absorption model

To motivate our modeling framework, we considered the absorption to follow a Weibull distribution. This type of absorption has been shown to capture the drug intake along the gastrointestinal tract and it is typically used whenever the absorption process cannot be described by zero-th order, first order process, or a combination of both.[20]

The estimated drug absorbed by a Weibull-type absorption is given by the following time-dependent formula[23]:

$$K_{\text{Weibull}}(t) = F\left(1 - e^{-(K_A(t-t_{\text{LastDose}}))^{\gamma}}\right) \quad (1)$$

where F is the absolute bioavailability, K_A is an absorption parameter, γ is a shaping factor that modulates the rate of absorption, and t_{LastDose} denotes the time of the last dose event.

In silico patient data

For simulating in silico patient data, we considered an NLME with a common system of ODEs as the structural model:

$$\begin{aligned} \frac{\text{dDepot}}{\text{d}t} &= -K_{\text{Weibull}}(t)\cdot \text{Depot}, \\ \frac{\text{dCentr}}{\text{d}t} &= K_{\text{Weibull}}(t)\cdot \text{Depot} - \frac{\text{CL}}{V}\text{Centr}, \end{aligned} \quad (2)$$

where Depot and Centr denote the depot and central compartment, respectively, CL is the apparent clearance, V is the apparent volume of distribution of the central compartment, and the drug exhibits a Weibull-type absorption process. Zero initial conditions were considered.

We assumed clearance exhibits interindividual variability, which we modeled by considering that the individual clearance for subject i is given by $\text{CL}_i = \text{CL}_{\text{pop}} e^{\eta_i}$where $\eta_i \sim \mathcal{N}(0, \sigma_{\text{IIV}})$. In this formula, CL_{pop} is the population apparent clearance, σ_{IIV} is the expected log-normal variability, and η_i is the individual random effect that we assume to follow a normal distribution with expectation 0 and standard deviation σ_{IIV}.

Let c_{ij} denote the observed concentrations for subject i and observation time t_j. We considered two situations:

Situation 1: drug concentration measurements exhibiting an additive error:

$$c_{ij} = \frac{\text{Centr}_{\text{ij}}}{V_i} + \varepsilon_{\text{ij}}, \text{where } \varepsilon_{\text{ij}} \sim \mathcal{N}(0, \sigma_{\text{Add}}),$$

Situation 2: drug concentration measurements exhibiting a proportional error:

$$c_{ij} = \frac{\text{Centr}_{\text{ij}}}{V_i}(1 + \varepsilon_{\text{ij}}), \text{where } \varepsilon_{\text{ij}} \sim \mathcal{N}(0, \sigma_{\text{Prop}}).$$

The parameters of the NLME model are in Table S1.

If a concentration measurement after including the error term is negative, that measurement is recorded as zero. This rarely happens given the small values of σ_{Add} and σ_{Prop} (see Table S1).

To ensure a better comparison with current state-of-the-art models,[7,8] the data used in this paper resemble the settings (number of patients and dose range), in which these particular models were trained. In consequence, the in silico patient data consisted of 800 patients that were randomly allocated into eight dose groups: 80 (once weekly), 100, 120, 140, 160, 180, 200, and 220 mg q.w. Weekly subcutaneous dosing was assumed. The PK data from these patients was assumed to be described by the NLME given by Equation 2. It was simulated for 70 days and recorded every 12 h.

We considered two different PK sampling scenarios: a very intensive sampling scenario that ensures proof-of-concept for different data-intensive ML algorithms, and a sparser, more realistic sampling strategy, representative of a dose escalation-expansion study.[24]

Sampling 1 (highly intensive sampling): Same sampling for all patients. Concentrations available every 12 h for 42 days.

Sampling 2 (less intensive sampling): Same sampling for all patients, only seven available measurements per cycle in three cycles (only after the first, third, and fifth dose administrations) at 0, 12, 24, 48, 72, 96, and 144 h after the dose was administered.

Test scenarios

To demonstrate the performance of the different methods, we generated different prediction scenarios which allowed us to evaluate the models' capabilities for several practically relevant questions. Two general schemes were

considered: (1) prediction of a new patients set consisting of 120 new patients that have not been seen by the models and (2) prediction on the same patients set that was used for training. In both sets, the time horizon was set to 70 days. The first 42 days used the specific training sampling. Thereafter we used sampling 1 (every 12 h) for prediction and visualization. For each set, the scenarios were defined as follows:

Extrapolation: No change was made in the dosage amount nor dosing schedule; doses were administered according to the same dosing schedule as in the training data.

No dose after: We stopped dosing after day 42. It meant that when using the same patient set, only one additional dose was given to the patient (at day 42) beyond those in the training set.

Dose missing: We assumed no dosed on the 49th day. After that, dosing according to the initial regimen proceeds.

New dosage amount: We changed the dosage amount for all the patients. Specifically, we tested using two new doses that were not used in the training data. The first one was in the range of the simulated data (i.e., 150 mg), and the second one was higher than the maximum dose used during training (i.e., 230 mg). No dose interruptions were considered in this scenario.

To verify the impact of sample size, we trained all models with a different number of patients, where the lowest number represents the most realistic scenario. When using the same patient set (except for the new dosages amount), we measured the performance after the training time span and for the case of a completely new patient sample, the metric was calculated over the complete time horizon. Specifically, as evaluation metrics, we reported the mean absolute percentage error (MAPE)[25] and the root squared mean prediction error (RSME)[26] for all tested scenarios:

$$\text{MAPE} = \frac{100\,\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - x_i}{y_i}\right| \tag{3}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - x_i\right)^2}{n}}$$

where n is the number of patients, and y_i and x_i are the observed/simulated and predicted values.

Pharmacokinetic-scientific machine learning

In our proposed PK-SciML architecture, we used a standard ODE system to describe the drug dynamics in a depot and a central compartment. This scheme allows us to capture the known physiology, such as apparent clearance and apparent volume, in the architecture itself which ensured that certain physiological principles were preserved.

$$\frac{\text{dDepot}}{\text{d}t} = -\zeta \cdot \text{Depot} \tag{4}$$

$$\frac{\text{dCentr}}{\text{d}t} = \zeta \cdot \text{Depot} - \frac{\text{CL}}{V}\text{Centr}$$

$$\text{Depot}(t=0) = 0; \text{Centr}(t=0) = 0$$

Our goal is to learn the absorption mechanism (ζ), the CL, and V (i.e., the parameters which define the ODE system). Hence, given a dose regimen, we can predict the concentration in the central compartment along the complete time horizon.

Further physiology that was embedded in the architecture revolves around initial conditions and dosing. For the dosing, we followed a similar approach to the one taken by Lu et al.,[7] wherein each time a dose is administered to the patient, the dosage is added to a compartment. Besides, we followed the approach adopted by a different work of Lu[19] wherein the initial conditions of the ODE are physiologically fixed to zero which enables the prediction of "new" patients and scenarios that have no data available as is often needed in applications.

A neural network φ_θ was trained to predict the unknown absorption rate ζ based on the data of the *Depot* compartment and the difference in time since the last dose was given to the patient (tsld). CL and V were learned as parameters in each backpropagation step. We implemented a potentially weighted mean squared error (MSE) as loss function to estimate the optimal values. However, to allow the model to focus on learning the differences in the concentration values for each timepoint, we added the MSE along the temporal axis and then averaged over the batch size as follows:

$$L(x) = \sum_{i=n}^{n}\sum_{j=1}^{t}\frac{\left(y_{i,j} - x_{i,j}\right)^2}{w_{i,j}} \tag{5}$$

where n is the number of patients, t is the number of time steps, $y_{i,j}$ and $x_{i,j}(\zeta, \text{CL}, V)$ are the real and predicted concentrations of patient I at time step j, respectively, $w_{i,j}$ are weights that can be chosen appropriately (i.e., 1 for normal MSE), corresponding to an additive error component, or $y_{i,j}$ to account for a proportional error model commonly used in PK modeling.

Additionally, to get the predicted concentration $x_{i,j}(\zeta, \text{CL}, V)$ we divided the second dimension of the output of the ODE system by the predicted volume. Figure 1 shows an overview of our framework.

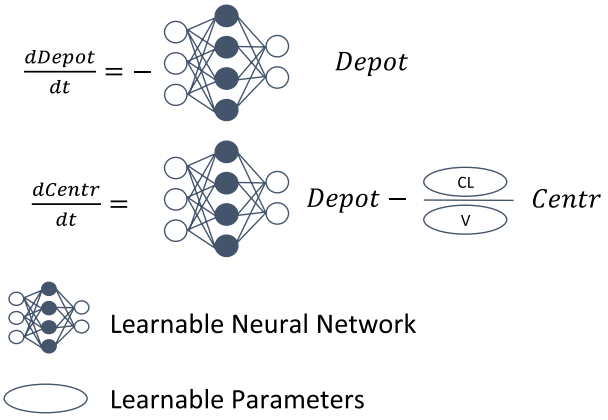

FIGURE 1 PK-SciML overview. We use a neural network to predict the absorption rate while at the same time we learn the clearance and the volume as individual parameters in each backward pass. PK-SciML, pharmacokinetic-scientific machine learning.

Model details

The neural network within our proposed PK-SciML consists of two fully connected layers with a *tanh* activation function in between and a SoftPlus[27] activation function at the end to avoid negative values. The first layer takes the concatenation of the *Depot* compartment information and the tsld as input and has 75 hidden neurons as output. The second layer contains a single output which represents the absorption rate. Although the weights θ of the neural network were initialized randomly, the CL and the V were initialized as 2.0 and 10.0, respectively. The model was trained with a batch size of 100 during 250 epochs using the Adaptive moment estimation (ADAM) optimizer[28] and a fixed learning rate of 0.01. Given the limited feasibility of performing a hyperparameter optimization for the Lu model[7] because of the long training time that is required, and to perform a fair comparison between the models, no hyperparameter optimization was run in this project for any model. Model implementation can be found in the Model Code of the Data S1.

Model comparison

We compared our approach against two state-of-the-art models for PK prediction, which are built based on neural ODE strategies[7] and integration of neural networks architectures.[8]

Lu et al.[7] proposed a framework which uses a GRU encoder to learn the initial conditions of a neural ODE system using the information about the dose cycle, dosage amount, PK information of the first cycle, time, and the time since the last dose. With the codified initial conditions, and the dosage amount sequence, an ODE solver was used to get the hidden state of the system which was then projected into the concentration using a decoder network. We followed the original implementation of Lu et al.[7] and trained the model using the RSME loss function (Equation 3), as well as the batch size and networks' hyperparameters as proposed by the authors. Unlike the original version, we increased the number of epochs from 30 to 100, to allow the model to better fit the training data. Without this increase, the model was found to predict underfitted concentrations during training. We used this approach in all the tables and figures as Lu et al.[7]

Bräm et al.[8] proposed a two-network architecture (i.e., dose-network and curve-network). The former consists of two branches which use information of concentration in the first cycle and the dosage amount sequence. Each branch has two long short-term memory (LSTM) units and a dense layer to get a representation which is then concatenated and projected to a 1 output using some dense layers. The curve-network also implements two LSTM units and two fully connected layers to get the representation from the complete concentration sequence. At the end, the outputs of both networks are added and used as the predicted concentration. We implemented the Bräm et al.[8] architecture in PyTorch[29] for a direct comparison between models. As proposed by the authors, the model was trained to predict the concentration of the next step. In that sense, the output of the network was concatenated to the concentration sequence until we could predict the complete trajectory. Different from the original implementation, we did not pre-train the dose-network using different learning rates and number of epochs. Instead, we directly trained the complete model in an end-to-end manner. Following the original implementation, the model was optimized using the MSE loss function and ADAM optimizer. Consistent with the Lu approach, we trained the model for 100 epochs and refer to this model in all tables and figures as Bräm et al.[8]

To go a step further, we developed a third model which does not need to use the PK information of the first cycle, as required by Lu et al.[7] and Bräm et al.[8] This modification was made to allow for predictions for new patients, because many practical applications of PK models require predictions for new patients and scenarios where first cycle information does not exist. More specifically, we removed: (1) the branch of the dose-network that uses the PK information and therefore the concatenation operation of the original version, and (2) the dense layer after the concatenation of the two branches of the dose-network. We train this model for 100 epochs using the same loss and optimizer as in Bräm et al.[8] In all tables and figures,

this adapted architecture is referred to as *Bräm w/o 1st PKC* which is short for "PK Cycle information."

RESULTS

PK-SciML extrapolates well

Figure 2 shows the mean drug concentration-time profiles, goodness-of-fit, and residual error plots of the extrapolation scenario using an additive error and dense sampling ($n=680$ patients). All compared models accurately predicted the mean concentration for each of the eight different dose groups during both training and extrapolation (Figure 2, left column). This finding is supported by the evaluation metrics (Table 1, Table S2), the goodness-of-fit and residual plots (middle and right column, respectively) where a close association between simulated and predicted values can be appreciated. Importantly, the residual plot illustrates a small drop in the prediction performance of the Lu et al.[7] approach ($t>50$), which is reflected by a mismatch in the last two cycles in the left plot. This suggests that this model will exhibit a decrease in performance for a long extrapolation task. This is likely a consequence of encoding the knowledge of the data into the initial conditions of the neural ODE, which then leads to a decay in their influence on the state of the system over time.

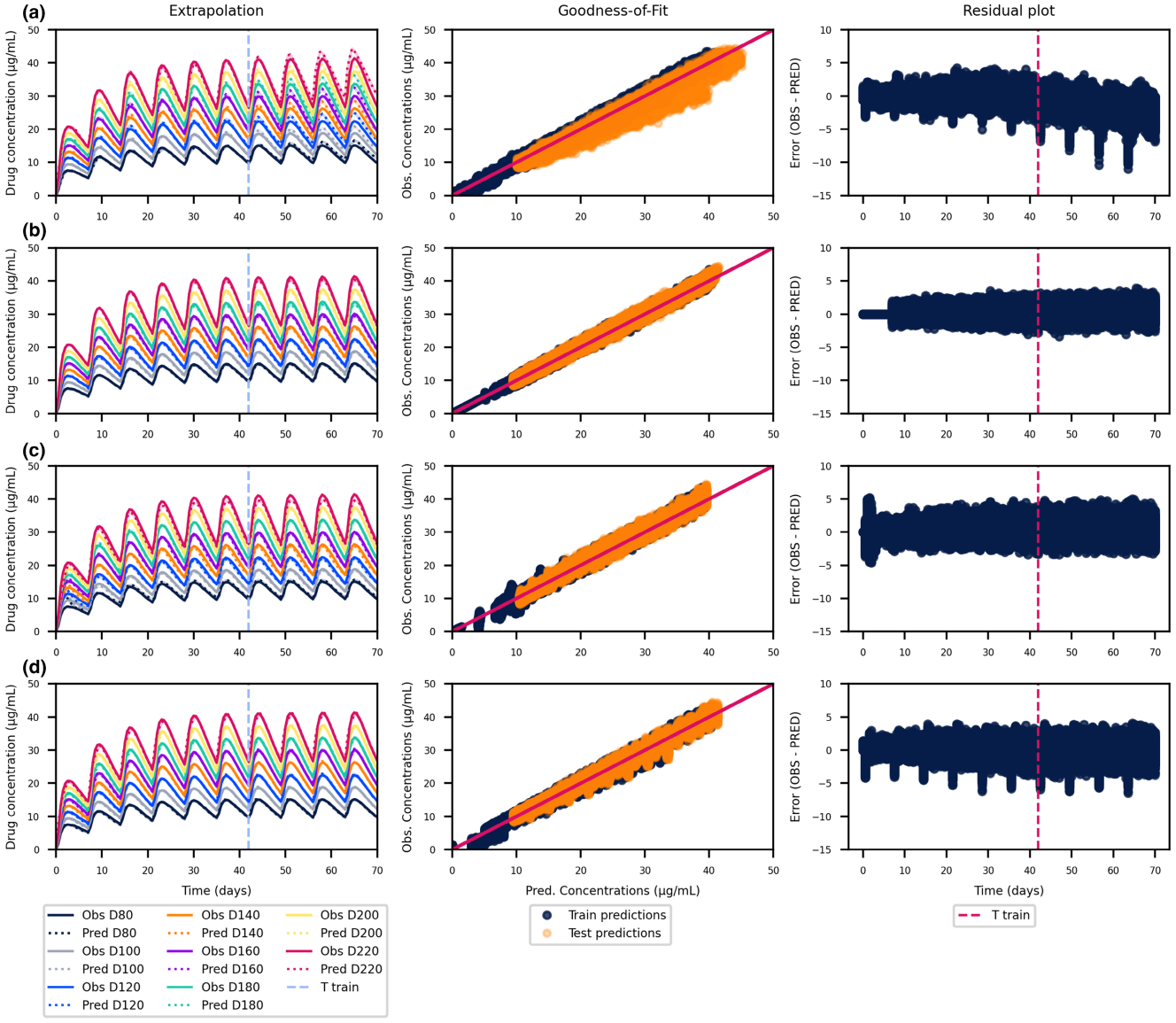

FIGURE 2 Extrapolation results when training the models using 680 patients on the additive error dataset with sparsity 1. (a) Lu, (b) Bräm, (c) *Bräm w/o 1st PKC*, (d) PK-SciML (ours). For each row we show from left to right the extrapolation results, the goodness-of-fit and the residual error plots. *Bräm w/o 1st PKC*, PK Cycle information; PK-SciML, pharmacokinetic-scientific machine learning, OBS, observed; PRED, predicted.

TABLE 1 Average MAPE for two different training scenarios on the same patients set.

	Additive error dataset - Sparsity 1						Proportional error dataset - Sparsity 2					
Model	Training patients	Extrapolation	No dose after	Dose missing	New dose (150 mg)	New dose (230 mg)	Training patients	Extrapolation	No dose after	Dose missing	New dose (150 mg)	New dose (230 mg)
Lu	680	7.8	133.0	23.3	7.5	6.1	680	16.1	22.3	17.0	13.8	14.0
	48	29.9	67.1	34.2	27.3	22.4	48	50.1	38.7	48.3	50.1	43.3
Bräm	680	**3.1**	113.7	6.1	**3.0**	**2.6**	680	62.7	50.5	59.0	61.2	61.2
	48	9.5	196.3	18.3	9.1	9.2	48	49.6	107.7	36.3	17.6	23.3
Bräm w/o 1st PKC	680	3.9	175.3	18.1	3.8	4.7	680	54.9	62.6	47.4	66.3	68.0
	48	29.2	347.4	50.3	15.0	42.4	48	32.2	366.9	61.2	14.8	36.2
PK-SciML	680	3.8	**15.9**	**4.9**	3.4	3.3	680	**8.8**	**11.5**	**9.2**	**8.7**	**8.7**
	48	**3.9**	**17.3**	**5.3**	**3.4**	**3.2**	48	**9.2**	**13.8**	**9.9**	**8.9**	**9.0**

Note: In bold, we show the best performance for each scenario. Metrics are reported after the training time.

Abbreviations: *Bräm w/o 1st PKC*, PK Cycle information; MAPE, mean absolute percentage error; PK-SciML, pharmacokinetic-scientific machine learning.

The results for a more realistic scenario (i.e., a dataset generated with a proportional error for 48 patients and less intensive sampling), are depicted in Figure 3. In contrast to our proposed approach, the models by Lu et al.[7] and Bräm et al.[8] cannot fit the training set with a small sample size. Specifically, the *Bräm w/o 1st PKC* architecture predicts similar concentrations for all doses close to the average of the first cycle concentration for all patients in the training set. This behavior confirms the importance of using data of the first PK cycle for this model, but also highlights the limited practical applicability. When assessing the ability to extrapolate using a different sampling rate, as opposed to our proposed method, the other models cannot adapt to the new scheme.

Our proposed model generates accurate predictions even when using few patients and extrapolates well to a different sampling scheme (Figure 3, left column). This interpretation is also supported by the goodness-of-fit and the prediction residual plots, which also indicate no substantial loss of prediction accuracy with longer extrapolation.

The findings for the no dose after and dose missing scenarios, as well as additional experiments with intermediate numbers of patients and sampling densities, are consistent with these findings (Figures S1 and S2). Although prediction errors for all models increase when the data are less abundant, our proposed architecture increases at a much slower rate compared to the other approaches and maintains a favorable performance across all tasks.

PK-SciML makes accurate predictions for new patients

Apart from our method, only the *Bräm w/o 1st PKC* approach[8] can be used in a predictive setting for entirely new patients. Like the extrapolation scenario, both models demonstrate an accurate prediction performance when using dense sampling for 680 patients (Figure 4, Table 2). However, in the more realistic less dense sampling situation with 48 patients, predictions by the *Bräm w/o 1st PKC* method[8] do not fit the data well, whereas our proposed approach was still accurate. As before, prediction errors generally decrease with the increasing number of patients (Table S2).

PK-SciML can predict concentrations for novel dosing regimens

Figure 5 presents the results for the three realistic scenarios, in which a novel dosing regimen is predicted. Models were trained with 48 patients and tested on the same set of patients at t greater than 49, akin to the extrapolation case. Although this task would also usually be in new patients, to allow for a better comparison, we have restricted it to those situations where all algorithms will produce a prediction. For the *Bräm w/o 1st PKC* model[8] we again observe a lack of fit to the training data and a high prediction error.

For the no dose scenario, the Lu et al.[7] predictions follow the expected behavior, whereas the Bräm et al.[8] approach predicts a steady-state at the last predicted value (Figure 5, left column).

In the dose missing scenario, where a dose is missing at $t = 49$, Lu et al.[7] and Bräm et al.[8] architectures under predict values at the time the dose is removed. As a result, the predicted concentration ranges for t greater than or equal to 56 are dissimilar to the real ones at times when patients are re-dosed. The predicted concentrations for the Lu et al.[7] approach are close to zero.

In contrast, our proposed SciML approach in all tested scenarios predicts the concentration with high accuracy.

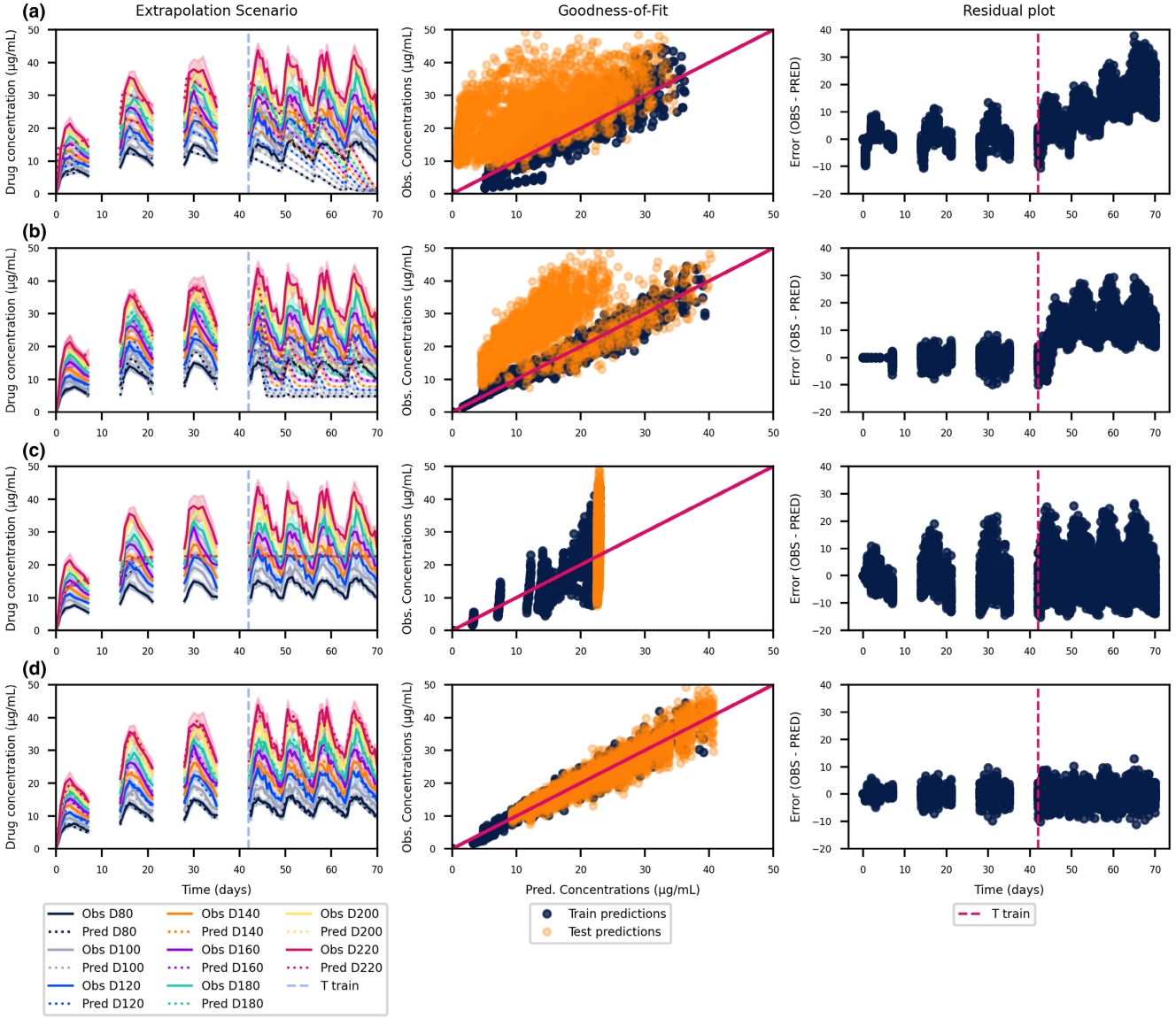

FIGURE 3 Extrapolation results when training the models using 48 patients on the proportional error dataset with the sparsity 2. (a) Lu, (b) Bräm, (c) *Bräm w/o 1st PKC*, (d) PK-SciML (ours). For each row we show from left to right the extrapolation results, the goodness-of-fit and the residual error plots. *Bräm w/o 1st PKC*, PK Cycle information; PK-SciML, pharmacokinetic-scientific machine learning, OBS, observed; PRED, predicted.

Thus, our model captures the mechanistic relationship between the dose and the concentration in the central compartment and can adapt the output to the dynamics of situations different to the one presented in the training set. The evaluation metrics for all tested scenarios are shown in Table S2.

To further support this point, the right column of Figure 5 shows the results when changing the dosage to a value that exceeds the maximum dose in the training data. Here, the Bräm et al.[8] as well as Lu et al.[7] architectures underestimate the concentration, whereas our PK-SciML approach accurately predicts concentrations that are close to real ones. Further supporting these findings, the point estimates of the apparent CL and apparent V made by our model are close to their original values used for generating the data (Table S1).

Influence of loss function on PK-SciML predictions

To investigate the real-world influence of increasing non-normal noise, we simulated an increasing amount of proportional error in the data and compare PK-SciML trained with the usual MSE versus a version trained with a proportional error loss (Equation 5). As expected, prediction

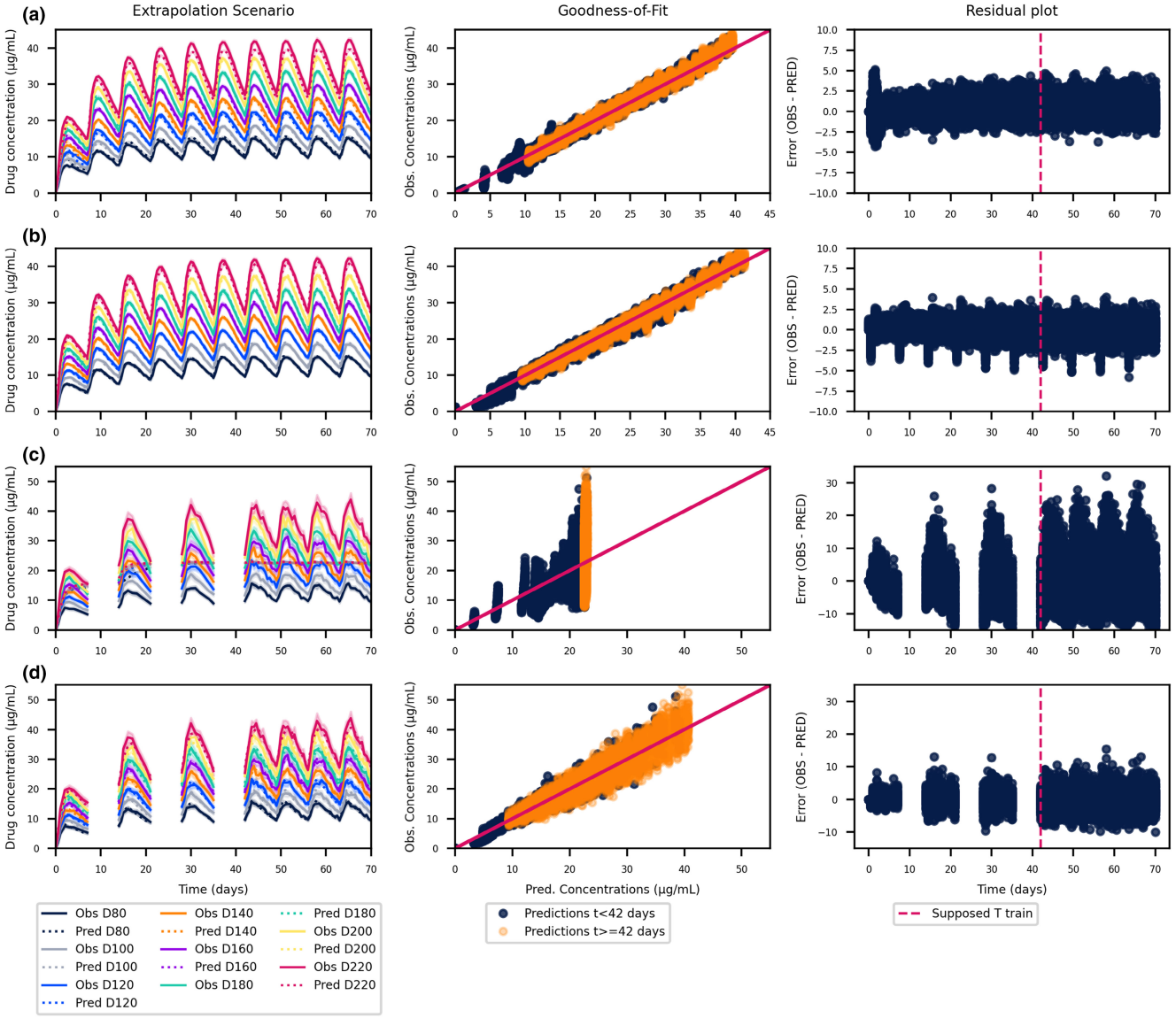

FIGURE 4 Prediction results on 96 new patients when (a), (b) training the models using 680 patients and sparsity 1, and (c, d) training the models using 48 patients and sparsity 2. (a, c) *Bräm w/o 1st PKC*, (b, d) PK-SciML (ours). For each row we illustrate from left to right the extrapolation results, the goodness-of-fit and the residual error plots. In the former, we keep the same training sparsity rate until $t = 42$ (i.e. training time). *Bräm w/o 1st PKC*, PK Cycle information; PK-SciML, pharmacokinetic-scientific machine learning, OBS, observed; PRED, predicted.

errors of the PK-SciML with proportional error loss were consistently lower than for the PK-SciML trained with the MSE as loss function, if the simulated amount of proportional error increases (Table S3). This shows that the structural model and loss function can be informed by principles known from PK modeling to benefit the reliability of the predictions.

DISCUSSION

Considering as a use case PK data with a complex absorption process, this work shows that incorporation of known PK mechanisms into ML models improve their performance. Our PK-SciML model was able to capture an unknown absorption mechanism, reflected by the highly accurate PK predictions, even when considering a realistic less intensive sampling scheme where the other models performed poorly. Moreover, the PK-SciML model allows the estimation of PK relevant parameters (i.e., apparent volume and apparent clearance), therefore allowing the interpretability of these values. This framework could be adapted in an ad hoc problem basis to learn an unknown mechanism present in the PK dynamics. To further exemplify this, in the Data S1, we show an example in which we adapted our architecture to capture an unknown and

TABLE 2 Average MAPE for two different training scenarios on the new patients set.

Model	Additive error dataset - Sparsity 1							Proportional error dataset - Sparsity 2						
	Training patients	Testing patients	Extrapolation	No dose after	Dose missing	New dose (150 mg)	New dose (230 mg)	Training patients	Testing patients	Extrapolation	No dose after	Dose missing	New dose (150 mg)	New dose (230 mg)
Bräm w/o 1st PKC	680	120	4.8	70.9	10.5	**3.9**	5.2	680	120	41.1	47.1	35.8	60.6	62.5
	48		36.0	174.0	44.7	18.3	37.0	48		34.4	274.5	55.5	18.7	32.2
PK-SciML	680	120	**4.4**	**8.6**	**4.7**	4.0	**3.9**	680	120	**9.2**	**11.1**	**9.5**	**8.9**	**8.9**
	48		**4.3**	**10.0**	**4.8**	**3.9**	**3.5**	48		**9.6**	**12.9**	**10.0**	**9.0**	**9.2**

Note: In bold, we show the best performance for each scenario. Metrics are reported in the complete time horizon.

Abbreviations: *Bräm w/o 1st PKC*, PK Cycle information; MAPE, mean absolute percentage error; PK-SciML, pharmacokinetic-scientific machine learning.

time-variant clearance while estimating simultaneously a linear absorption rate and volume (Figure S3). We observed a good fit and note that for different applications, the hyperparameters of our model should be chosen appropriately.

Another important observation is that without loss of performance, our proposed architecture also allows us to deal with one of the most relevant scenarios in PK modeling, namely predictions for new situations and patients with no available concentration data. Other prediction models[7,8] must be initialized with at least the concentration values of the drug during the first cycle we want to test, which is unavailable for new patients. In addition, our model is sufficiently robust such that: (1) it removes the difficulty of testing new doses, which is not possible for the other evaluated models because they must be initialized for a new dose group, even if the same patients are used, (2) it adapts to different extrapolation samples, and (3) it captures the effect that the dose has on the concentration, allowing for highly accurate predictions of dosing regimens which differ from the ones used in the training data. Last, all results are obtained without extensive hyperparameter optimization.

An important limitation of our model is that it generates population trajectories without considering the patients' individual parameters on the drug concentration. Addressing this limitation is planned for future work, and different methods can be used to consider patient specific parameters in a ML framework.[30–32] Another limitation, is that due to the lack of publicly available data, we could only use in silico data in this study. Moreover, the performance comparison must be considered with the fact in mind that the PK-SciML framework used the known true structure, which enabled the superior performance. However, PK models are always developed by considering known physiological structure and thus we do not see this aspect as a main limitation of our study.

Whereas our last test scenario resembles a very large phase I study in the number of patients and sampling scheme, it remains to be investigated how the performance of this method changes in settings with more sparse sampling, such as phase II/III studies or smaller phase I studies. A possible direction for improvement is uncertainty quantification of learned parameters (e.g., via variational inference).

We note that because the model has predefined structure to capture different PK processes, it is susceptible to the same problems of misspecification, as known for population PK models – with the additional benefit of replacing the misspecified part of the model with a neural network. This work shows the potential

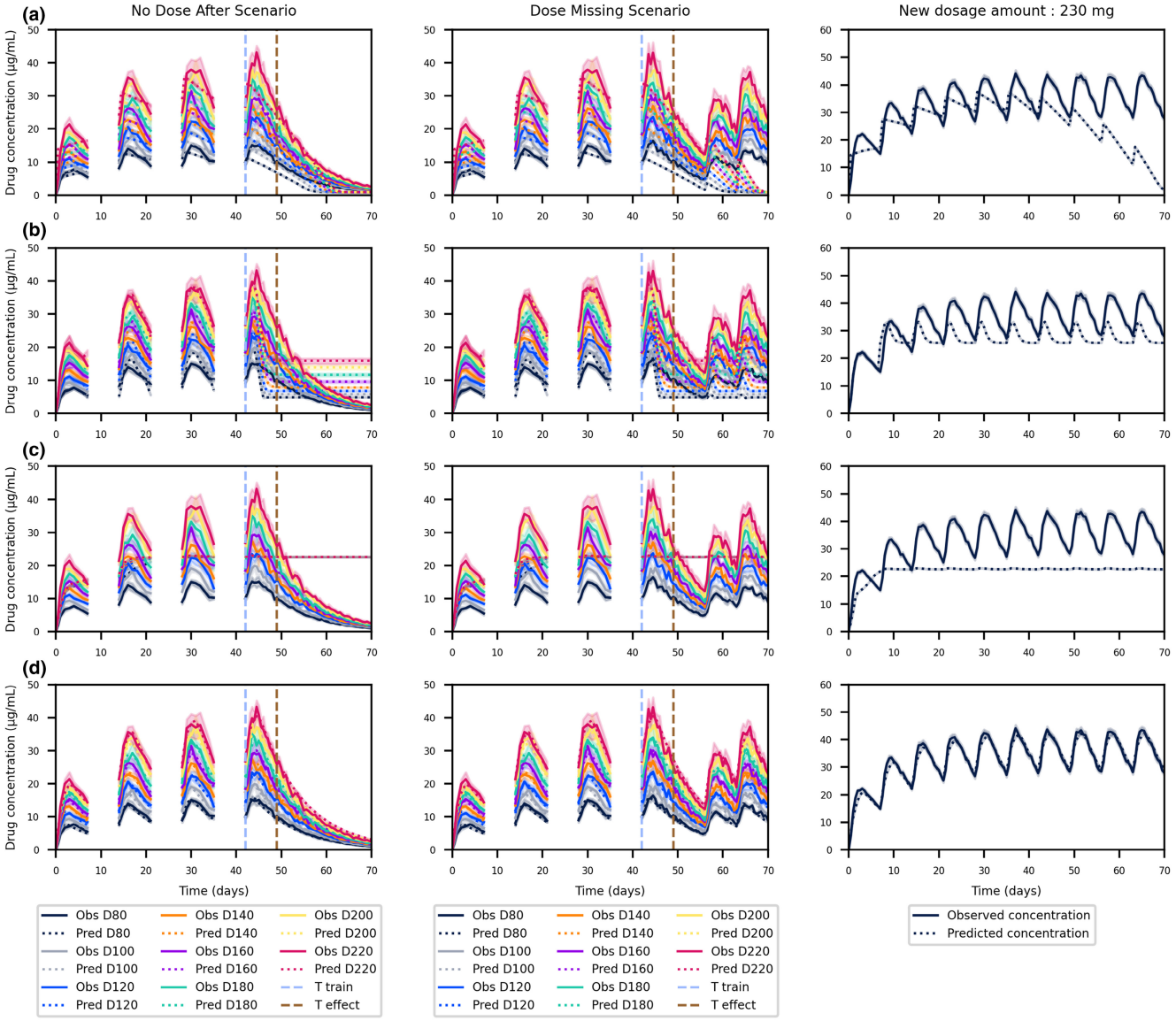

FIGURE 5 Prediction results on the same patients set using different scenarios. (a) Lu, (b) Bräm, (c) *Bräm w/o 1st PKC*, (d) PK-SciML (ours). For each row we depict from left to right the results of no dose after, dose missing and new dosage (230 µg) amount scenarios. All the models were trained using 48 patients on the proportional error dataset with sparsity 2. For the first two scenarios, we keep the same training sparsity rate until $t = 42$ (i.e. training time). *Bräm w/o 1st PKC*, PK Cycle information; PK-SciML, pharmacokinetic-scientific machine learning, OBS, observed; PRED, predicted.

of the development of hybrid models, which are at the same time more flexible than traditional models and more interpretable than traditional ML models, relying on an expert human teaming up with the algorithm to improve upon what each could have achieved in isolation.[7,8]

AUTHOR CONTRIBUTIONS

D.V., A.V.P.B., S.M., H.F., and S.S., wrote the manuscript. H.F. and S.S. designed the research. D.V performed the research. D.V. and A.V.P.B. analyzed the data.

ACKNOWLEDGMENTS

Medical writing support was provided by Wesley Wayman, PhD, an employee of AbbVie.

FUNDING INFORMATION

This study was sponsored by AbbVie, Inc.

CONFLICT OF INTEREST STATEMENT

A.V.P.B., S.M., and S.S. are employees of AbbVie and may hold AbbVie stock. Fraunhofer SCAI (employees: D.V. and H.F.) received funding from AbbVie for this study. AbbVie contributed to the study design, research, and

interpretation of data, and the writing, reviewing, and approving of the publication.

ORCID

Diego Valderrama https://orcid.org/0000-0002-3450-0359
Ana Victoria Ponce-Bobadilla https://orcid.org/0000-0002-0959-4058
Sven Stodtmann https://orcid.org/0000-0002-7986-4447

REFERENCES

1. Bräm DS, Nahum U, Atkinson A, Koch G, Pfister M. Evaluation of machine learning methods for covariate data imputation in pharmacometrics. *CPT Pharmacometrics Syst Pharmacol*. 2022;11:1638-1648.
2. Sibieude E, Khandelwal A, Hesthaven JS, Girard P, Terranova N. Fast screening of covariates in population models empowered by machine learning. *J Pharmacokinet Pharmacodyn*. 2021;48:597-609.
3. Janssen A, Hoogendoorn M, Cnossen MH, Mathôt RAA, for the OPTI-CLOT Study Group and SYMPHONY Consortium. Application of SHAP values for inferring the optimal functional form of covariates in pharmacokinetic modeling. *CPT Pharmacomet Syst Pharmacol*. 2022;11:1100-1110.
4. Zhu X, Zhang M, Wen Y, Shang D. Machine learning advances the integration of covariates in population pharmacokinetic models: Valproic acid as an example. *Front Pharmacol*. 2022;13:994665.
5. Marzano L, Darwich AS, Tendler S, et al. A novel analytical framework for risk stratification of real-world data using machine learning: a small cell lung cancer study. *Clin Transl Sci*. 2022;15:2437-2447.
6. Qian Z, Zame W, Fleuren L, Elbers P, van der Schaar M. Integrating expert ODEs into neural ODEs: pharmacology and disease progression. *Adv Neural Inform Process Syst*. 2021;34:11364-11383.
7. Lu J, Deng K, Zhang X, Liu G, Guan Y. Neural-ODE for pharmacokinetics modeling and its advantage to alternative machine learning models in predicting new dosing regimens. *iScience*. 2021;24:102804.
8. Bräm DS, Parrott N, Hutchinson L, Steiert B. Introduction of an artificial neural network-based method for concentration-time predictions. *CPT Pharmacometrics Syst Pharmacol*. 2022;11:745-754.
9. Janssen A, Leebeek FWG, Cnossen MH, Mathôt RAA, for the OPTI-CLOT study group and SYMPHONY consortium. Deep compartment models: a deep learning approach for the reliable prediction of time-series data in pharmacokinetic modeling. *CPT: Pharmacometrics Syst Pharmacol*. 2022;11:934-945.
10. Keutzer L, You H, Farnoud A, et al. Machine learning and pharmacometrics for prediction of pharmacokinetic data: differences, similarities and challenges illustrated with rifampicin. *Pharmaceutics*. 2022;14:1530.
11. Terranova N, Venkatakrishnan K, Benincosa LJ. Application of machine learning in translational medicine: current status and future opportunities. *AAPS J*. 2021;23:74.
12. Janssen A, Bennis FC, Mathôt RAA. Adoption of machine learning in pharmacometrics: an overview of recent implementations and their considerations. *Pharmaceutics*. 2022;14:1814.
13. Chen RT, Rubanova Y, Bettencourt J, Duvenaud DK. Neural ordinary differential equations. *Adv Neural Inform Process Syst*. 2018;31:6571-6583.
14. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. 2015. International Conference on Medical Image Computing and Computer-Assisted Intervention: 234–241.
15. Baker N, Alexander F, Bremer T, et al. Workshop Report on Basic Research Needs for Scientific Machine Learning: Core Technologies for Artificial Intelligence. 2019.
16. Rackauckas C, Ma Y, Martensen J, et al. Universal differential equations for scientific machine learning. *arXiv*. 2020. Preprint arXiv 2001.04385.
17. Ramadhan A, Marshall J, Souza A, et al. Capturing missing physics in climate model parameterizations using neural differential equations. *arXiv*. 2020. Preprint arXiv.2010.12559.
18. Dandekar R, Rackauckas C, Barbastathis G. A machine learning-aided global diagnostic and comparative tool to assess effect of quarantine control in COVID-19 spread. *Patterns (N Y)*. 2020;1:100145.
19. Lu J, Bender B, Jin JY, Guan Y. Deep learning prediction of patient response time course from early data via neural-pharmacokinetic/pharmacodynamic modelling. *Nat Mach Intell*. 2021;3:696-704.
20. Zhou H. Pharmacokinetic strategies in deciphering atypical drug absorption profiles. *J Clin Pharmacol*. 2003;43:211-227.
21. Gomeni R, Bressolle-Gomeni F. Modeling complex pharmacokinetics of long-acting injectable products using convolution-based models with nonparametric input functions. *J Clin Pharmacol*. 2021;61:1081-1095.
22. Greenblatt DJ, Shader RI. Drug absorption rate: a critical component of bioequivalence assessment in psychopharmacology. *J Clin Pharmacol*. 1987;27:85-86.
23. Bonate PL, Howard DR. *Pharmacokinetics in Drug Development*. Springer; 2011.
24. Araujo DV, Oliva M, Li K, Fazelzad R, Liu ZA, Siu LL. Contemporary dose-escalation methods for early phase studies in the immunotherapeutics era. *Eur J Cancer*. 2021;158:85-98.
25. De Myttenaere A, Golden B, Le Grand B, Rossi F. Mean absolute percentage error for regression models. *Neurocomputing*. 2016;192:38-48.
26. Barnston AG. Correspondence among the correlation, RMSE, and Heidke forecast verification measures; refinement of the Heidke score. *Weather Forecastng*. 1992;7:699-709.
27. Zheng H, Yang Z, Liu W, Liang J, Li Y. Improving deep neural networks using softplus units. 2015. International Joint Conference on Neural Networks (IJCNN) 2015;1–4.
28. Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv*. 2014. Preprint arXiv 1412.6980.
29. Paszke A, Gross S, Massa F, et al. PyTorch: an imperative style, high-performance deep learning library. *Advances in neural information processing systems*. 2019;32:8026-8037.
30. Mandel F, Ghosh RP, Barnett I. Neural networks for clustered and longitudinal data using mixed effects models. *Biometrics*. 2021;79:711-721.

31. Nazarovs J, Chakraborty R, Tasneeyapant S, Ravi SN, Singh V. Mixed effects neural ODE: a Variational approximation for analyzing the dynamics of panel data. *arXiv*. 2022. Preprint arXiv 2202.09463.
32. Simchoni G, Rosset S. Integrating random effects in deep neural networks. *arXiv*. 2022. Preprint arXiv 2206.03314.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Valderrama D, Ponce-Bobadilla AV, Mensing S, Fröhlich H, Stodtmann S. Integrating machine learning with pharmacokinetic models: Benefits of scientific machine learning in adding neural networks components to existing PK models. *CPT Pharmacometrics Syst Pharmacol*. 2024;13:41-53. doi:10.1002/psp4.13054